

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2002, 11:00:11 ; Search time 29.95 Seconds
(without alignments)
467.289 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 126
Sequence: 1 AQSVPGGDIQTQPGTKIVFN.....EWFQDGMVRRKNLPIEYNP 126

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	11.9	16	AAW31324	Dictyocaulus vivip
2	9	7.1	10	AAW31325	Dictyocaulus vivip
3	8	6.3	8	AAW31323	Dictyocaulus vivip
4	8	6.3	20	AAW31322	Dictyocaulus vivip
5	8	6.3	454	AAW33365	Sequence of purine
6	7	5.6	73	AAO01320	Human polypeptide
7	7	5.6	142	AAW80296	Human protein SEQ
8	7	5.6	151	AAO12550	Human polypeptide
9	7	5.6	154	AAG74054	Human colon cancer
10	7	5.6	155	AAG26775	Zea mays protein f
11	7	5.6	160	AAW80302	Human protein SEQ

12	7	5.6	161	22	AAW92552	Human protein sequ
13	7	5.6	175	21	AAG26274	Arabidopsis thalia
14	7	5.6	175	21	AAG46875	Arabidopsis thalia
15	7	5.6	178	22	AAG89942	C glutamic prote
16	7	5.6	180	22	AAO06983	Human polypeptide
17	7	5.6	184	21	AAG46874	Arabidopsis thalia
18	7	5.6	187	21	AAG26273	Arabidopsis thalia
19	7	5.6	506	22	AAW78344	Human protein SEQ
20	7	5.6	509	22	AAU34300	Staphylococcus aur
21	7	5.6	510	21	AAW68778	Amino acid sequenc
22	7	5.6	517	22	AAU37244	Staphylococcus aur
23	7	5.6	527	22	AAW79328	Human protein SEQ
24	7	5.6	978	22	AAU33960	Staphylococcus aur
25	7	5.6	1001	22	AAU37093	Staphylococcus aur
26	7	5.6	1066	17	AAU91300	TATA box binding p
27	7	5.6	1066	19	AAW33632	Yeast transcriptio
28	7	5.6	1076	22	ABG20205	Novel human diagno
29	7	5.6	1551	22	AAW66597	Candida albicans A
30	6	4.8	12	22	AAU25777	Breast cancer-asso
31	6	4.8	15	21	AAW79881	Human papillomavir
32	6	4.8	29	14	AAW30690	Synthetic LamB4 si
33	6	4.8	29	14	AAW30176	Synthetic LamB3 si
34	6	4.8	29	22	ABB39402	Peptide #6908 enco
35	6	4.8	29	22	ABB24184	Protein #6183 enco
36	6	4.8	29	22	AAW60082	Human brain expres
37	6	4.8	29	22	AAW72691	Human bone marrow
38	6	4.8	29	22	AAW19685	Peptide #6119 enco
39	6	4.8	29	22	AAW32922	Peptide #6959 enco
40	6	4.8	34	22	AAU18940	Novel lung cancer
41	6	4.8	34	22	AAU17952	Novel human respir
42	6	4.8	46	21	AAW08061	Arabidopsis thalia
43	6	4.8	46	21	AAG61624	Arabidopsis thalia
44	6	4.8	46	22	AAW88344	Human immune/haema
45	6	4.8	52	22	AAU56623	Propionibacterium

ALIGNMENTS

RESULT	1
AAW31324	
ID	AAW31324 standard; peptide; 16 AA.
XX	
AC	AAW31324;
XX	
DT	13-MAR-1998 (first entry)
XX	
DE	Dictyocaulus viviparus DV18 peptide fragment 4.
XX	
KW	Immunogenic protein; DV18; lungworm; vaccine; immunity;
KW	dictyocauliasis; cattle; immunoassay.
XX	
OS	Dictyocaulus viviparus.
XX	
PN	EP785253-A1.
XX	
PD	23-JUL-1997.
XX	
PF	27-DEC-1996; 96EP-0120947.
XX	
PR	19-JAN-1996; 96DE-4001754.
XX	
PA	(FARH) HOECHST AG.
XX	
PI	Hofmann J, Schmid K;
XX	
DR	WPI; 1997-365928/34.
XX	
PT	Immunogenic Dictyocaulus viviparus lungworm protein - for use in
PT	vaccines and immunoassays
XX	
PS	Example 10; Page 6; 17pp; German.
XX	

CC This sequence represents a peptide fragment of the immunogenic protein
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid
CC can be joined to the amino acids Asp, Asn, Ser, Gly or Arg. DV18 can be
CC used in vaccines for immunising cattle against dictyocauliasis and in an
CC ELISA immunoassay for determining DV18-specific antibodies in the blood
CC of cattle.
XX
SQ Sequence 16 AA;

Query Match 11.9%; Score 15; DB 18; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.8e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 FRREWFQDGMVRRK 118
|||||
Db 2 frrewfgdgmvrk 16

RESULT 2
AAW31325
ID AAW31325 standard; peptide; 10 AA.
XX
AC AAW31325;

DT 13-MAR-1998 (first entry)
XX
DE Dictyocaulus viviparus DV18 peptide fragment 5.
XX
DE Immunogenic protein; DV18; lungworm; vaccine; immunity;
KW dictyocauliasis; cattle; immunoassay.
XX
OS Dictyocaulus viviparus.

XX EP785253-A1.
XX
PD 23-JUL-1997.
XX
PF 27-DEC-1996; 96EP-0120947.
XX
PR 19-JAN-1996; 96DE-4001754.
XX
PA (FARH) HOECHST AG.

XX Hofmann J, Schmid K;
XX WPI; 1997-365928/34.
XX
XX Immunogenic Dictyocaulus viviparus lungworm protein - for use in
PT vaccines and immunoassays
XX
PS Example 10; Page 6; 17pp; German.

XX This sequence represents a peptide fragment of the immunogenic protein
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid
CC can be joined to the amino acids Asp, Asn, Ser, Gly or Arg. DV18 can be
CC used in vaccines for immunising cattle against dictyocauliasis and in an
CC ELISA immunoassay for determining DV18-specific antibodies in the blood
CC of cattle.
XX
SQ Sequence 10 AA;

Query Match 7.1%; Score 9; DB 18; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 VFNAPYDDK 26
|||||
Db 2 vfnapyddk 10

RESULT 3

AAW31323
ID AAW31323 standard; peptide; 8 AA.
XX
AC AAW31323;
XX
DT 13-MAR-1998 (first entry)
XX
DE Dictyocaulus viviparus DV18 peptide fragment 3.
XX
KW Immunogenic protein; DV18; lungworm; vaccine; immunity;
KW dictyocauliasis; cattle; immunoassay.
XX
OS Dictyocaulus viviparus.

XX EP785253-A1.
XX
PD 23-JUL-1997.
XX
PF 27-DEC-1996; 96EP-0120947.
XX
PR 19-JAN-1996; 96DE-4001754.

XX (FARH) HOECHST AG.
XX
PI Hofmann J, Schmid K;
XX
DR WPI; 1997-365928/34.

XX Immunogenic Dictyocaulus viviparus lungworm protein - for use in
PT vaccines and immunoassays
XX
PS Example 10; Page 6; 17pp; German.

XX This sequence represents a peptide fragment of the immunogenic protein
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid
CC can be joined to the amino acids Asp, Asn, Ser, Gly or Arg. DV18 can be
CC used in vaccines for immunising cattle against dictyocauliasis and in an
CC ELISA immunoassay for determining DV18-specific antibodies in the blood
CC of cattle.
XX
SQ Sequence 8 AA;

Query Match 6.3%; Score 8; DB 18; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 NLPYEYNP 126
|||||
Db 1 nlpieynp 8

RESULT 4
AAW31322
ID AAW31322 standard; peptide; 20 AA.
XX
AC AAW31322;

DT 13-MAR-1998 (first entry)
XX
DE Dictyocaulus viviparus DV18 peptide fragment 2.
XX
KW Immunogenic protein; DV18; lungworm; vaccine; immunity;
KW dictyocauliasis; cattle; immunoassay.

XX Dictyocaulus viviparus.

XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /label= unknown
XX
PN EP785253-A1.
XX

PD 23-JUL-1997.
XX
XX 27-DEC-1996; 96EP-0120947.
PF
XX 19-JAN-1996; 96DE-4001754.
PR
XX (FARH) HOECHST AG.
PA
XX Hofmann J, Schmid K;
PI
XX WPI; 1997-365928/34.
DR
XX Immunogenic Dictyocaulus viviparus lungworm protein - for use in
PT vaccines and immunoassays
PT
XX
PS Example 10; Page 6; 17pp; German.
XX
CC This sequence represents a peptide fragment of the immunogenic protein
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid
CC can be joined to the amino acids Asp, Asn, Ser, Gly or Arg. DV18 can be
CC used in vaccines for immunising cattle against dictyocauliasis and in an
CC ELISA immunoassay for determining DV18-specific antibodies in the blood
CC of cattle.
XX
SQ Sequence 20 AA;

Query Match 6.3%; Score 8; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GVLDPKEA 67
Db | | | | | | | |
9 gvl dpkea 16

RESULT 5
AAR33365
ID AAR33365 standard; Protein; 454 AA.
XX
AC AAR33365;
XX
DT 30-JUN-1993 (first entry)
XX
DE Sequence of purine-rich repeat (GA repeat) binding protien
DE (GABP) subunit alpha.
XX
KW GA binding protein; cis-regulatory element;
KW VP16 mediated induction.
XX
OS Mus musculus.
XX
PN WO9304166-A.
XX
PD 04-MAR-1993.
XX
PF 17-AUG-1992; 92WO-US06748.
XX
PR 16-AUG-1991; 91US-0746032.
XX
PA (CARN-) CARNEGIE INST WASHINGTON.
XX
PI Lamarco KL, Mc Knight SL, Thompson CC;
XX
DR WPI; 1993-093998/11.
DR N-PSDB; AAQ37480.
XX
PT DNA encoding GA binding protein sub-unit - allows investigation
PT of sub-unit sequence motif functions, for control of rapid cell
PT division e.g. in cancer
XX
PS Disclosure; Fig 2A; 68pp; English.
XX

CC A cis-regulatory element required for virion associated protein VP16
CC mediated induction of herpes simplex virus 1 (HSV1) immediate early
CC (IE) genes consists of three imperfect repeats of the purine-rich
CC hexanucleotide 5'-CGGAAR-3'. A protein complex capable of avid
CC interaction with the purine-rich repeats (GA repeats) has been
CC identified in soluble preparations of rat liver nucleic. This GA
CC binding protein (GABP) consists of two separable subunits.
CC Applicants have isolated cDNA clones encoding both subunits of GABP
CC and have revealed that one (GABP alpha) is related to the Ets
CC transforming protein, while the other (GABP beta) contains a
CC series of 33-amino acid repeats related in sequence to a variety
CC of proteins.
XX
SQ Sequence 454 AA;

Query Match 6.3%; Score 8; DB 14; Length 454;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IKVINSSA 38
Db | | | | | | | |
284 ikvinssa 291

RESULT 6
AAO01320
ID AAO01320 standard; Protein; 73 AA.
XX
AC AAO01320;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 15212.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI81251.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 15212; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 73 AA;

Query Match 5.6%; Score 7; DB 22; Length 73;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 PDGAARQ 103
|||||||
Db 24 pdgaarq 30

RESULT 7
AAM80296
ID AAM80296 standard; Protein; 142 AA.
XX
AC AAM80296;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 3948.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX

OS Homo sapiens.

XX WO200157190-A2.

PN
XX
PD 09-AUG-2001.

XX
PF 05-FEB-2001; 2001WO-US04098.

XX
PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang 2W;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK53429.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX

PS Claim 20; Page 6210-6211; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 142 AA;

Query Match 5.6%; Score 7; DB 22; Length 142;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AVLLAVS 73
|||||||
Db 2 avllavs 8

RESULT 8
AAO12550
ID AAO12550 standard; Protein; 151 AA.
XX
AC AAO12550;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 26442.

DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX

OS Homo sapiens.

XX WO200164835-A2.

PN
XX
PD 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI92481.

XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX

PS Claim 20; SEQ ID NO 26442; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

XX Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 151 AA;

Query Match 5.6%; Score 7; DB 22; Length 151;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	97	PDGAARQ 103	
Db	40	pdgaarq 46	
RESULT 9			
AAG74054			
ID	AAG74054	standard; Protein; 154 AA.	
XX			
AC	AAG74054;		
XX			
DT	03-SEP-2001	(first entry)	
XX			
DE	Human colon cancer antigen protein SEQ ID NO:4818.		
XX			
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; chromosome 13.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200122920-A2.		
XX			
PD	05-APR-2001.		
XX			
PF	28-SEP-2000; 2000WO-US26524.		
XX			
PR	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
DR	WPI; 2001-235357/24.		
DR	N-PSDB; AAH33485.		
XX			
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
PT	useful for preventing, diagnosing and/or treating colorectal cancers -		
XX			
PS	Claim 11; Page 6603-6604; 9803pp; English.		
XX			
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patients own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated Ps,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		
CC	to express the proteins. N and P can be used in the prevention, diagnosis		
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204		
CC	and AAH77789 represent sequences used in the exemplification of the		
CC	present invention.		
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were		
CC	missing at time of publication, meaning no sequences are present for		
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.		
XX			
SQ	Sequence 154 AA;		

Query Match 5.6%; Score 7; DB 22; Length 154;
Best Local Similarity 100.0%; Pred. NO. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	65	KEAVLLA 71	
Db	103	keavlla 109	
RESULT 10			
AAG26775			
ID	AAG26775	standard; Protein; 155 AA.	
XX			
AC	AAG26775;		
XX			
DT	17-OCT-2000	(first entry)	
XX			
DE	Zea mays protein fragment SEQ ID NO: 31357.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence; corn.		
XX			
OS	Zea mays subsp. mays.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137502.		
PR	07-JUN-1999; 99US-0137724.		
PR	08-JUN-1999; 99US-0138094.		
PR	10-JUN-1999; 99US-0138540.		
PR	10-JUN-1999; 99US-0138847.		
PR	14-JUN-1999; 99US-0139119.		
PR	16-JUN-1999; 99US-0139452.		
PR	16-JUN-1999; 99US-0139453.		

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.6%; Score 7; DB 21; Length 155;
Best Local Similarity 100.0%; Pred.No.19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 PDGAARQ 103

|||||

Db 12 pdgaarq 18

RESULT 11

AAM80302
ID AAM80302 standard; Protein; 160 AA.
XX
AC AAM80302;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3960.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;

WPI; 2001-476283/51.
N-PSDB; AAK53435.

Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -

Claim 20; Page 477; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.
Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
(AAM80020) are omitted as the relevant pages from the sequence listing
were missing at the time of publication.

Sequence 160 AA;

Query Match 5.6%; Score 7; DB 22; Length 160;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AVLLAVS 73
|||||||
Db 20 avllavs 26

RESULT 12

AAB92552

ID AAB92552 standard; Protein; 161 AA.

XX
AC AAB92552;

XX
DT 26-JUN-2001 (first entry)

XX
DE Human protein sequence SEQ ID NO:10737.

XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX
OS Homo sapiens.

XX
PN EP1074617-A2.

XX
PD 07-FEB-2001.

XX
PF 28-JUL-2000; 2000EP-0116126.

XX
PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX
PA (HELI-) HELIX RES INST.

XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX
DR WPI; 2001-318749/34.

XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX
PS Claim 8; SEQ ID 10737; 2537pp + CD ROM; English.

XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAK03166 to AAK13628 and
CC AAK13633 to AAK18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAK13629 to AAK13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 161 AA;

Query Match 5.6%; Score 7; DB 22; Length 161;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 KRLGVDP 57
|||||||
Db 120 krlgvdp 126

RESULT 13
AAG26274
ID AAG26274 standard; Protein; 175 AA.
XX
AC AAG26274;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 30670.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.6%; Score 7; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 AVLLAVS 73
| | | | |
Db 54 avllavs 60

RESULT 14
AAG46875
ID AAG46875 standard; Protein; 175 AA.
XX
AC AAG46875;
XX

DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 59019.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.6%; Score 7; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AVLLAVS 73
|||||
Db 54 avllavs 60

RESULT 15
AAG89942
ID AAG89942 standard; Protein; 178 AA.

XX AC AAG89942;

XX DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 3696.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis.

OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65161.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 3696; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 178 AA;

Query Match 5.6%; Score 7; DB 22; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AVLLAVS 73
|
Db 71 avllavs 77

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2002, 11:00:36 ; Search time 12.89 Seconds
(without alignments)
238.761 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 126
Sequence: 1 AQSVPBGDIQTQPGTKIVFN.....EWFQDGMVRRKNLPIEYNP 126

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	23	18.3	127	4	US-08-786-455B-14	Sequence 14, Appl
2	10	7.9	14	4	US-08-786-455B-4	Sequence 4, Appli
3	9	7.1	9	4	US-08-786-455B-5	Sequence 5, Appli
4	8	6.3	20	4	US-08-786-455B-2	Sequence 2, Appli
5	7	5.6	1066	2	US-08-308-818-1	Sequence 1, Appli
6	7	5.6	1551	4	US-09-425-665-2	Sequence 2, Appli
7	7	5.6	1551	4	US-09-685-668-2	Sequence 2, Appli
8	6	4.8	7	4	US-08-786-455B-3	Sequence 3, Appli
9	6	4.8	34	3	US-08-658-136-34	Sequence 34, Appl
10	6	4.8	71	4	US-08-928-213B-23	Sequence 23, Appl
11	6	4.8	106	2	US-08-785-065-11	Sequence 11, Appl
12	6	4.8	128	3	US-08-906-769-143	Sequence 143, App
13	6	4.8	128	3	US-08-906-616-143	Sequence 143, App
14	6	4.8	128	4	US-08-639-075A-143	Sequence 143, App
15	6	4.8	128	4	US-09-012-431-143	Sequence 143, App
16	6	4.8	128	4	US-09-012-692-143	Sequence 143, App
17	6	4.8	128	4	US-08-906-613-143	Sequence 143, App
18	6	4.8	209	4	US-09-196-293-15	Sequence 15, Appl
19	6	4.8	210	1	US-08-158-353-3	Sequence 3, Appli
20	6	4.8	210	4	US-08-209-603E-15	Sequence 15, Appl
21	6	4.8	210	4	US-08-235-836C-30	Sequence 30, Appl
22	6	4.8	252	4	US-09-199-637A-176	Sequence 176, App
23	6	4.8	264	1	US-08-107-042-2	Sequence 2, Appli
24	6	4.8	273	1	US-08-215-928A-2	Sequence 2, Appli
25	6	4.8	273	1	US-08-446-920-2	Sequence 2, Appli
26	6	4.8	273	4	US-08-928-213B-9	Sequence 9, Appli
27	6	4.8	309	2	US-08-701-191A-9	Sequence 9, Appli

28	6	4.8	347	1	US-08-446-920-10	Sequence 10, Appl
29	6	4.8	349	3	US-08-762-500-77	Sequence 77, Appl
30	6	4.8	386	1	US-08-134-012-3	Sequence 3, Appli
31	6	4.8	386	1	US-08-520-519-3	Sequence 3, Appli
32	6	4.8	466	4	US-08-235-836C-107	Sequence 107, App
33	6	4.8	466	4	US-08-235-836C-110	Sequence 110, App
34	6	4.8	476	4	US-09-346-408-12	Sequence 12, Appl
35	6	4.8	482	3	US-09-135-639-2	Sequence 2, Appli
36	6	4.8	500	4	US-08-960-190A-25	Sequence 25, Appl
37	6	4.8	528	4	US-08-928-213B-8	Sequence 8, Appli
38	6	4.8	568	4	US-08-637-823B-27	Sequence 27, Appl
39	6	4.8	572	2	US-09-032-315-7	Sequence 7, Appli
40	6	4.8	572	2	US-08-993-318A-7	Sequence 7, Appli
41	6	4.8	572	4	US-09-399-886-7	Sequence 7, Appli
42	6	4.8	572	4	US-09-396-260-7	Sequence 7, Appli
43	6	4.8	572	4	US-09-576-281-7	Sequence 7, Appli
44	6	4.8	584	4	US-08-637-823B-28	Sequence 28, Appl
45	6	4.8	588	4	US-08-235-836C-122	Sequence 122, App

ALIGNMENTS

RESULT 1
US-08-786-455B-14
; Sequence 14, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheirich
; TITLE OF INVENTION: DICTYOCAULUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-786-455B-14

Query Match 18.3%; Score 23; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 QFREWFQDGMVRRKNLPIEYN 125
Db 103 QFREWFQDGMVRRKNLPIEYN 125

RESULT 2
US-08-786-455B-4
; Sequence 4, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheirich
; TITLE OF INVENTION: DICTYOCAULUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-786-455B-4

Query Match 7.9%; Score 10; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 FRREWFQDGDG 113
Db 1 FRREWFQDGDG 10

RESULT 3
US-08-786-455B-5
; Sequence 5, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheirich
; TITLE OF INVENTION: DICTYOCAULUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-786-455B-5

Query Match 7.1%; Score 9; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 VFNAPYDDK 26
Db 1 VFNAPYDDK 9

RESULT 4
US-08-786-455B-2
; Sequence 2, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheirich
; TITLE OF INVENTION: DICTYOCAULUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.

```

;
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-786-455B-2

Query Match 6.3%; Score 8; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GVLDPKEA 67
Db 9 GVLDPKEA 16
|||||

RESULT 5
US-08-308-818-1
; Sequence 1, Application US/08308818
; Patent No. 5847077
; GENERAL INFORMATION:
; APPLICANT: Green, Michael R
; APPLICANT: Reese, Joseph C
; TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein
; TITLE OF INVENTION: Complex Critical for Expression of Fungal Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,818
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0342/0A404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-52707700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: S. cerevisiae
; IMMEDIATE SOURCE:
; CLONE: TAF-145
; US-08-308-818-1
```

```

Query Match 5.6%; Score 7; DB 2; Length 1066;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 PGTKIVF 19
Db 504 PGTKIVF 510
|||||

RESULT 6
US-09-425-665-2
; Sequence 2, Application US/09425665
; Patent No. 6174705
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Richard O.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: ARO1
; FILE REFERENCE: GM10245
; CURRENT APPLICATION NUMBER: US/09/425,665
; CURRENT FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1551
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-425-665-2

Query Match 5.6%; Score 7; DB 4; Length 1551;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 VLLAVSC 74
Db 1453 VLLAVSC 1459
|||||

RESULT 7
US-09-685-668-2
; Sequence 2, Application US/09685668
; Patent No. 6346405
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Richard O.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: ARO1
; FILE REFERENCE: GM10245
; CURRENT APPLICATION NUMBER: US/09/685,668
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 09/425,665
; PRIOR FILING DATE: 1999-10-2209
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1551
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-685-668-2

Query Match 5.6%; Score 7; DB 4; Length 1551;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 VLLAVSC 74
Db 1453 VLLAVSC 1459
|||||

RESULT 8
US-08-786-455B-3
; Sequence 3, Application US/08786455B
; Patent No. 6193971
```

```
;
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; TITLE OF INVENTION: DICTYOCALUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-786-455B-3

Query Match 4.8%; Score 6; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 LPIEYN 125
Db 1 LPIEYN 6

RESULT 9
US-08-658-136-34
; Sequence 34, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
```

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-658-136-34

Query Match 4.8%; Score 6; DB 3; Length 34;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QSVPPG 7
Db 18 QSVPPG 23

RESULT 10
US-08-928-213B-23
; Sequence 23, Application US/08928213B
; Patent No. 6238905
; GENERAL INFORMATION:
; APPLICANT: McHenry, Charles S.
; SEVILLE, Mark
; CULL, Millard G.
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III
; NUMBER OF SEQUENCES: 195
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,213B
; FILING DATE: 12-Sep-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: MacKnight, Kamrin T.
; REGISTRATION NUMBER: 38,230
; REFERENCE/DOCKET NUMBER: ENZYCO-02550
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-705-8410
; TELEFAX: 415-397-8338
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
```

```

;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-08-928-213B-23

      Query Match      4.8%; Score 6; DB 4; Length 71;
      Best Local Similarity 100.0%; Pred. No. 39;
      Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      56 DPPCGV 61
Db      1 DPPCGV 6

RESULT 11
US-08-785-065-11
; Sequence 11, Application US/08785065
; Patent No. 5814451
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL SUBUNITS OF NADH DEHYDROGENASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,065
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0187 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 224
US-08-785-065-11

      Query Match      4.8%; Score 6; DB 2; Length 106;
      Best Local Similarity 100.0%; Pred. No. 57;
      Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      51 KRLGVD 56
Db      8 KRLGVD 13

RESULT 12
```

```

US-08-906-769-143
; Sequence 143, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-906-769-143

      Query Match      4.8%; Score 6; DB 3; Length 128;
      Best Local Similarity 100.0%; Pred. No. 67;
      Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      15 TKIVFN 20
Db      42 TKIVFN 47

RESULT 13
US-08-906-616-143
; Sequence 143, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-906-616-143

Query Match 4.8%; Score 6; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TKIVEN 20
Db 42 TKIVEN 47

RESULT 14
US-08-639-075A-143
; Sequence 143, Application US/08639075A
; Patent No. 6150125
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,075A
; FILING DATE: 24-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
```

```

; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-639-075A-143

Query Match 4.8%; Score 6; DB 4; Length 128;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TKIVEN 20
Db 42 TKIVEN 47

RESULT 15
US-09-012-431-143
; Sequence 143, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,431
; FILING DATE: 23-Jan-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-09-012-431-143
```


Query Match 4.8%; Score 6; DB 4; Length 128;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 TKIVFN 20
| | | | |
Db 42 TKIVFN 47

Search completed: September 24, 2002, 11:03:14
Job time: 158 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2002, 11:01:01 ; Search time 15.96 Seconds
(without alignments)
758.600 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 126
Sequence: 1 AQSVPGGDIQTQPGTKIVN.....EFQGDGMVRRKNLPIEYNP 126

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	100.0	127	2 T16183	hypothetical prote
2	101	80.2	127	2 T21640	hypothetical prote
3	101	80.2	127	2 G88145	protein F58A6.8 [i
4	101	80.2	127	2 A88165	protein ZK1248.6 [
5	101	80.2	127	2 G88686	protein msp-19 [im
6	101	80.2	127	2 C88688	protein msp-113 [i
7	101	80.2	127	2 H88688	protein msp-59 [im
8	101	80.2	127	2 B88689	protein msp-65 [im
9	101	80.2	127	2 C88689	protein msp-51 [im
10	101	80.2	127	2 H88792	protein K07F5.1 [i
11	101	80.2	127	2 H88146	protein C34F11.4 [
12	101	80.2	127	2 F88134	protein msp-40 [im
13	101	80.2	127	2 F88138	protein MSP-31 [im
14	101	80.2	127	2 D88164	protein msp-142 [i
15	100	79.4	127	2 A88139	protein ZK546.6 [i
16	93	73.8	190	2 T16687	major sperm protei
17	90	71.4	127	2 T24885	hypothetical prote
18	86	68.3	127	2 F88801	protein C04G2.4 [i
19	71	56.3	127	2 C88164	protein K05F1.7 [i
20	68	54.0	127	2 T16684	major sperm protei
21	59	46.8	127	2 A88683	protein C09B9.6 [i
22	53	42.1	127	2 F88146	protein C34F11.6 [
23	52	41.3	77	2 F88165	protein ZK1248.4 [
24	51	40.5	133	2 T27902	hypothetical prote
25	39	31.0	127	2 T23486	hypothetical prote
26	23	18.3	127	2 A45944	major sperm protei
27	23	18.3	127	2 A45528	major sperm protei
28	23	18.3	127	2 B45528	major sperm protei
29	19	15.1	95	2 T34500	hypothetical prote

30	11	8.7	484	2 T26393	hypothetical prote
31	9	7.1	99	2 T31606	hypothetical prote
32	8	6.3	454	1 A48146	GA-binding protein
33	8	6.3	454	1 A40858	GA-binding protein
34	7	5.6	37	2 T29808	hypothetical prote
35	7	5.6	167	2 C70866	hypothetical prote
36	7	5.6	178	2 S29119	hypothetical prote
37	7	5.6	203	2 AH1434	AA3-600 quinol oxi
38	7	5.6	203	2 AH1806	AA3-600 quinol oxi
39	7	5.6	231	2 B83031	conserved hypothet
40	7	5.6	249	2 A70786	probable Cobs - My
41	7	5.6	278	2 H87663	hypothetical prote
42	7	5.6	282	2 E83086	conserved hypothet
43	7	5.6	286	2 C70953	probable lpq2 prot
44	7	5.6	347	2 T06329	symbiotic ammonium
45	7	5.6	391	2 A69372	ammonium transport

ALIGNMENTS

RESULT 1
T16183
hypothetical protein F26G1.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 26-May-2000
C:Accession: T16183
R:Chisoe, S.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid F26G1.
A:Reference number: Z18472
A:Accession: T16183
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-127 <CHI>
A:Cross-references: EMBL:U23519; NID:g746524; PID:g746531; PIDN:AAC46807.1; CESP:F26G
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F26G1.7
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 100.0%; Score 126; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 2.5e-127;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||
Db 2 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRTTVEWNTPTDGAARQERREWFQGDGMVRRKNL 120
|||||
Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRTTVEWNTPTDGAARQERREWFQGDGMVRRKNL 121
QY 121 PIEYNP 126
|||||
Db 122 PIEYNP 127

RESULT 2
T21640
hypothetical protein F32B6.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
C:Accession: T21640; T24884
R:Basham, V.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19453
A:Accession: T21640
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-127 <WIL>
A:Cross-references: EMBL:Z81074; PIDN:CAB03037.1; GSPDB:GN00022; CESP:F32B6.6

A; Experimental source: clone F32B6
R; Swinburne, J.
submitted to the EMBL Data Library, October 1996
A; Reference number: Z19947
A; Accession: T24884
A; Status: preliminary; translated from GB/EMBL/DBBJ
A; Molecule type: DNA
A; Residues: 1-127 <WI2>
A; Cross-references: EMBL:Z81122; PIDN:CAB03361.1; GSPDB:GN00022; CESP:T13F2.10
A; Experimental source: clone T13F2
C; Genetics:
A; Gene: CESP:F32B6.6; CESP:T13F2.10
A; Map position: 4
C; Superfamily: *Caenorhabditis elegans* major sperm protein

```

Query Match      80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||
2 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||

61 VLPKKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
|||||
62 VLPKKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102
|||||

```

RESULT 3
G88145
protein F58A6.8 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88145
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88145
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_II; PIDN:AA96204.1; PID:g12555857; GSPDB:GN00020; CESP:F58A6.8
C:Genetics:
A:Gene: F58A6.8
A:Map position: 2
C:Superfamily: Caenorhabditis elegans major sperm protein

	Query Match	80.2%;	Score 101;	DB 2;	Length 127;
	Best Local Similarity	100.0%;	Pred. No. 1.5e-100;		
	Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AQSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP	PCG	60	
Db	2	AQSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP	PCG	61	
QY	61	VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA	101		
Db	62	VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA	102		

```

RESULT      4
A881165
protein ZK1248.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A881165
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

```

A;Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C-
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: A88165
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_II; PID:g862495; GSPDB:GN00020; CESP:ZK1248.6
A;Note: similar to *C. elegans* major sperm protein
C;Genetics:
A;Gene: ZK1248.6
A;Map position: 2
C;Superfamily: *Caenorhabditis elegans* major sperm protein

	Query Match	80.2%	Score 101;	DB 2;	Length 127;
	Best Local Similarity	100.0%;	Pred. No. 1.5e-100;		
	Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AQSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGLVDPPCG	60		
Db	2	AQSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGLVDPPCG	61		
QY	61	VLDPKEAVLLAVSCDAFAFCQEDTNNDRITVEWTNTPDGAA	101		
Db	62	VLDPKEAVLLAVSCDAFAFCQEDTNNDRITVEWTNTPDGAA	102		

RESULT 5
G88686
protein msp-19 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88686
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investi
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/P
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 21
A:Accession: G88686
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAC26926.1; PID:g3329619; GSPDB:GN000022
C:Genetics:
A:Gene: msp-19
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

	Query Match	80.2%;	Score 101;	DB 2;	Length 127;
	Best Local Similarity	100.0%;	Pred. No. 1.5e-100;		
	Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AQSVPPGDIQTQPGTKIVENAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP	PCG	60	
Db	2	AQSVPPGDIQTQPGTKIVENAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDP	PCG	61	
QY	61	VLDPKEAVLLAVSCDAFAGQEDTNNDRITVEWTNTPDGAA	101		
Db	62	VLDPKEAVLLAVSCDAFAGQEDTNNDRITVEWTNTPDGAA	102		

```

RESULT      6
C88688
protein msp-113 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: C88688
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

```

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: C88688
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB42255.1; PID:g1825633; GSPDB:GN00022
C;Genetics:
A;Gene: msp-113
A;Map position: 4
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 7
H88688
protein msp-59 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: H88688
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: H88688
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB42253.1; PID:g1825631; GSPDB:GN00022
C;Genetics:
A;Gene: msp-59
A;Map position: 4
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 8
H88689
protein msp-65 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: H88689
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological

A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_eli
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: B88689
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB42256.1; PID:g1825634; GSPDB:GN00022
C;Genetics:
A;Gene: msp-65
A;Map position: 4
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 9
C88689
protein msp-51 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: C88689
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_eli
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: C88689
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB42254.1; PID:g1825632; GSPDB:GN00022
C;Genetics:
A;Gene: msp-51
A;Map position: 4
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 10
H88792
protein K07F5.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: H88792
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A;Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: H88134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA94282.1; PID:g3878316; GSPDB:GN000022; CESP:K07F5.
C:Genetics:
A:Gene: K07F5.1
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 101
|||||
Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 102
|||||

RESULT 11

H88146

protein C34F11.4 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: H88146

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: H88146

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-127 <STO>

A:Cross-references: GB:chr_II; PIDN:AA85761.1; PID:g1166627; GSPDB:GN000020; CESP:C34F11.

A:Note: similar to major sperm protein

C:Genetics:

A:Gene: C34F11.4

A:Map position: 2

C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;

Best Local Similarity 100.0%; Pred. No. 1.5e-100;

Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 101
|||||

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 102
|||||

RESULT 12

E88134

protein msp-40 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: E88134

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_--
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: E88134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_II; PIDN:AA933398.1; PID:g1203940; GSPDB:GN000020; CESP:C33F
A:Note: C33F10.9
C:Genetics:
A:Gene: msp-40
A:Map position: 2
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 101
|||||

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 102
|||||

RESULT 13

F88138

protein MSP-31 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: F88138

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_--

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: F88138

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-127 <STO>

A:Cross-references: GB:chr_II; PIDN:AA83175.1; PID:g1109821; GSPDB:GN000020; CESP:R05F

A:Note: R05F9.13

C:Genetics:

A:Gene: MSP-31

A:Map position: 2

C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;

Best Local Similarity 100.0%; Pred. No. 1.5e-100;

Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 101
|||||

Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIITVEWTNTPDGAA 102
|||||

RESULT 14

D88164

protein msp-142 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: D88164

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: D88164
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_II; PID:g868174; GSPDB:GN00020; CESP:K05F1.2
A;Note: K05F1.2
C;Genetics:
A;Gene: msp-142
A;Map position: 2
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 80.2%; Score 101; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPCCG 60
|||
Db 2 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPCCG 61

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNDTRITVEWTNTPDGAA 101
|||
Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNDTRITVEWTNTPDGAA 102

RESULT 15

A88139
protein ZK546.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: A88139
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88139
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:chr_II; PID:g868205; GSPDB:GN00020; CESP:ZK546.6
A;Note: Major sperm protein
C;Genetics:
A;Gene: ZK546.6
A;Map position: 2
C;Superfamily: Caenorhabditis elegans major sperm protein

Query Match 79.4%; Score 100; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.8e-99;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPCCGV 61
|||
Db 3 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPCCGV 62

QY 62 LDPKEAVLLAVSCDAFAFGQEDTNDTRITVEWTNTPDGAA 101
|||
Db 63 LDPKEAVLLAVSCDAFAFGQEDTNDTRITVEWTNTPDGAA 102

Search completed: September 24, 2002, 11:03:36
Job time: 155 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2002, 11:02:21 ; Search time 11.85 Seconds
(without alignments)
411.702 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 126
Sequence: 1 AQSVPFGDIQTQPGTKIVFN.....EWFQDGMVRRKNLPIEYNP 126

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	80.2	126	MS31_CAEEL	P53017 caenorhabdi
2	93	73.8	190	MS32_CAEEL	P53018 caenorhabdi
3	68	54.0	126	MS33_CAEEL	P53019 caenorhabdi
4	56	44.4	126	MS10_CAEEL	P05634 caenorhabdi
5	53	42.1	126	MS56_CAEEL	P05635 caenorhabdi
6	39	31.0	126	MS38_CAEEL	P53020 caenorhabdi
7	23	18.3	126	MSP1_ASCSU	P27439 ascaris suu
8	23	18.3	126	MSP1_ONCVO	P13262 onchocerca
9	23	18.3	126	MSP2_ASCSU	P27440 ascaris suu
10	23	18.3	126	MSP2_ONCVO	P13263 onchocerca
11	19	15.1	125	MSP1_GLORO	P53021 globodera r
12	19	15.1	125	MSP2_GLORO	P53022 globodera r
13	19	15.1	125	MSP3_GLORO	P53023 globodera r
14	8	6.3	265	KNH1_CANGA	O74684 candida gla
15	8	6.3	454	GABA_HUMAN	Q06546 homo sapien
16	8	6.3	454	GABA_MOUSE	Q00422 mus musculu
17	7	5.6	178	R20K_CLOPA	P23161 clostridium
18	7	5.6	249	COBS_MYCTU	Q10397 mycobacteri
19	7	5.6	511	NEK3_MOUSE	Q9r0a5 mus musculu
20	7	5.6	1066	T145_YEAST	P46677 saccharomyc
21	7	5.6	1106	ACLY_CAEEL	P53585 caenorhabdi
22	6	4.8	64	RL1_STRLA	O87734 streptomyce
23	6	4.8	102	HCC2_CRYCO	Q01238 crypthecodi
24	6	4.8	105	NIPM_BOVIN	Q02379 bos taurus
25	6	4.8	115	MOTI_BOVIN	O62820 bos taurus
26	6	4.8	121	KDGL_ECOLI	P00556 escherichia
27	6	4.8	129	VMOR_BPMU	P23848 bacterioph
28	6	4.8	135	Y360_METJA	Q57806 methanococc
29	6	4.8	137	ATPE_PICAB	O47036 picea abies
30	6	4.8	144	MARR_SALTY	Q56069 salmonella
31	6	4.8	150	SP0A_BACCE	P52930 bacillus ce
32	6	4.8	177	GPIX_HUMAN	P14770 homo sapien
33	6	4.8	182	RRF_SYNY3	P74456 synechocyst

RESULT 1					ALIGNMENTS				
ID	MS31_CAEEL	STANDARD;	PRT;	126 AA.					
AC	P53017;								
DT	01-OCT-1996 (Rel. 34, Created)								Q9ka66 bacillus ha
DT	01-OCT-1996 (Rel. 34, Last sequence update)								P52933 bacillus pu
DT	16-OCT-2001 (Rel. 40, Last annotation update)								Q28679 oryctolagus
DE	Major sperm protein 31/40/142 (MSP).								P48803 bos taurus
GN	(MSP-31 OR R05F9.13) AND (MSP-40 OR C33F10.9) AND (MSP-51 OR ZK354.5)								Q07337 borrelia bu
GN	AND (MSP-59 OR ZK354.11) AND (MSP-65 OR ZK354.1) AND (MSP-113 OR								P52931 bacillus ci
GN	ZK354.4) AND (MSP-142 OR K05F1.2) AND C34F11.4 AND F58A6.8 AND K07F5.1								P52932 bacillus me
GN	AND ZK1248.6.								P34460 caenorhabdi
OS	Caenorhabditis elegans.								P52928 bacillus an
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;								Q92q57 rickettsia
OC	Rhabditidae; Peloderinae; Caenorhabditis.								P97054 rhodobacter
OX	NCBI_Taxid=6239;								P33281 magnaporthe
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Hallsworth K.;								
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBBJ databases.								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Waterston R.;								
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBBJ databases.								
RN	[3]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Wohldmann P.;								
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBBJ databases.								
RN	[4]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Bentley D.;								
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBBJ databases.								
RN	[5]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Waterston R.;								
RL	Submitted (APR-1996) to the EMBL/GenBank/DBBJ databases.								
RN	[6]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Hembry C.;								
RL	Submitted (MAR-1996) to the EMBL/GenBank/DBBJ databases.								
RN	[7]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=BRISTOL N2;								
RA	Latreille P.;								
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBBJ databases.								
RN	[8]								
RP	SEQUENCE FROM N.A. (MSP-51; MSP-59; MSP-65 AND MSP-113).								
RC	STRAIN=BRISTOL N2;								
RA	Johnson D., Wamsley P., Bradshaw H.;								
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBBJ databases.								

CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U29244; AAC71087.1; -
DR EMBL; U29377; AAA68711.1; -
DR EMBL; U46753; AAA85761.1; -
DR EMBL; U49830; AAK31477.1; -
DR EMBL; U53339; AAA96204.1; -
DR EMBL; U41533; AAA83175.1; -
DR EMBL; Z70284; CAA94282.1; -
DR EMBL; U88172; AAB42253.1; -
DR EMBL; U88172; AAB42254.1; -
DR EMBL; U88172; AAB42255.1; -
DR EMBL; U88172; AAB42256.1; -
DR HSSP; P27439; 3MSP.
DR WormPep; C33F10.9; CE02806.
DR WormPep; C34F11.4; CE02806.
DR WormPep; F58A6.8; CE02806.
DR WormPep; K05F1.2; CE02806.
DR WormPep; K07F5.1; CE02806.
DR WormPep; R05F9.13; CE02806.
DR WormPep; ZK1248.6; CE02806.
DR WormPep; ZK354.1; CE09978.
DR WormPep; ZK354.4; CE09978.
DR WormPep; ZK354.5; CE09978.
DR WormPep; ZK354.11; CE09978.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0
FT MOD_RES 1
SQ SEQUENCE 126 AA; 14078 MW; 0F069631D8559AB7 CRC64;

Query Match 80.2%; Score 101; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 3e-101;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 60
Db 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCG 60

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101

RESULT 2
MS32_CAEEL
ID MS32_CAEEL STANDARD; PRT; 190 AA.
AC P53018;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 32 (MSP).
GN MSP-32 OR R05F9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hallsworth K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U41533; AAA83168.1; -
DR HSSP; P27439; 1MSP.
DR WormPep; R05F9.3; CE04805.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Cytoskeleton; Sperm; Multigene family.
SQ SEQUENCE 190 AA; 21304 MW; 939C5B8FD79C54BA CRC64;

Query Match 73.8%; Score 93; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 1.7e-92;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 IQTPGCTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCGVLDPKPEAV 68
Db 73 IQTPGCTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPVPCGVLDPKPEAV 132

QY 69 LLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 133 LLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 165

RESULT 3
MS33_CAEEL
ID MS33_CAEEL STANDARD; PRT; 126 AA.
AC P53019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 33 (MSP).
GN MSP-33 OR R05F9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hallsworth K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----


```
-----
CC EMBL; U41533; AAA83165.1; -.
DR HSSP; P27439; 3MSP.
DR WormPep; R05F9.8; CE04811.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 126 AA; 14112 MW; 0F0C9C31D25F9AB7 CRC64;

Query Match 54.0%; Score 68; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.9e-66;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
Db 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60

QY 61 VLDPKAV 68
Db 61 VLDPKAV 68

RESULT 4
MS10_CAEEL STANDARD; PRT; 126 AA.
ID MS10_CAEEL
AC P05634;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 10 (MSP).
GN MSP-10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=84191131; PubMed=6325882;
RA Klass M.R., Kinsley S., Lopez L.C.;
RT "Isolation and characterization of a sperm-specific gene family in
the nematode Caenorhabditis elegans.";
RL Mol. Cell. Biol. 4:529-537(1984).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; K02617; AAA28115.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 126 AA; 14093 MW; 835CA9DBFD80D5D9 CRC64;

Query Match 44.4%; Score 56; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 7.1e-53;
```

```
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 KTTNMKRLGVDPGCVLDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAA 101
Db 46 KTTNMKRLGVDPGCVLDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAA 101

RESULT 5
MS56_CAEEL STANDARD; PRT; 126 AA.
ID MS56_CAEEL
AC P05635;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 56 (MSP).
GN MSP-56.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=84191131; PubMed=6325882;
RA Klass M.R., Kinsley S., Lopez L.C.;
RT "Isolation and characterization of a sperm-specific gene family in
the nematode Caenorhabditis elegans.";
RL Mol. Cell. Biol. 4:529-537(1984).
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; K02618; AAA28116.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 126 AA; 14063 MW; 82F7B36A4D80C5C8 CRC64;

Query Match 42.1%; Score 53; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.2e-49;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 NMKRLGVDPGCVLDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAA 101
Db 49 NMKRLGVDPGCVLDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAA 101

RESULT 6
MS38_CAEEL STANDARD; PRT; 126 AA.
ID MS38_CAEEL
AC P53020;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major sperm protein 38 (MSP).
GN MSP-38 OR K08F4.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Hembry C.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
 CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
 CC -!- TISSUE SPECIFICITY: SPERM.
 CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
 CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z68879; CAA93089.1; -.
 DR HSSP; P27439; 3MSP.
 DR Wormpep; K08F4.8; CE06156.
 DR InterPro; IPR000535; MSP_domain.
 DR Pfam; PF00635; MSP_domain; 1.
 KW Cytoskeleton; Acetylation; Sperm; Multigene family.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 SQ SEQUENCE 126 AA; 14064 MW; 98F8888DFD8BB7A49 CRC64;

 Query Match 31.0%; Score 39; DB 1; Length 126;
 Best Local Similarity 100.0%; Pred. No. 1.3e-34;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 63 DPKEAVLLAVSCDAFAFGQEDTNDRTITVEWNTPTDGA 101
 |||||
 DB 63 DPKEAVLLAVSCDAFAFGQEDTNDRTITVEWNTPTDGA 101
 |||||

 RESULT 7
 MSP1_ASCSU STANDARD; PRT; 126 AA.
 ID MSP1_ASCSU
 AC P27439; P27441;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Major sperm protein, isoform alpha (Alpha-MSP).
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Sperm;
 RX MEDLINE=92407055; PubMed=1527183;
 RA King K.L., Stewart M., Roberts T.M., Seavy M.;
 RT "Structure and macromolecular assembly of two isoforms of the major
 RT sperm protein (MSP) from the amoeboid sperm of the nematode, Ascaris
 RT suum.";
 RL J. Cell Sci. 101:847-857(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87031211; PubMed=3770294;
 RA Bennett K.L., Ward S.;
 RT "Neither a germ line-specific nor several somatically expressed genes
 RT are lost or rearranged during embryonic chromatin diminution in the
 RT nematode Ascaris lumbricoides var. suum.";
 RL Dev. Biol. 118:141-147(1986).
 RN [3]
 RP SEQUENCE FROM N.A.

RA Bullock T.L., Parathasathy G., King K.L., Kent M.L., Roberts T.M.,
 RA Stewart M.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=97070381; PubMed=8913307;
 RA Bullock T.L., Roberts T.M., Stewart M.;
 RT "2.5-A resolution crystal structure of the motile major sperm protein
 RT (MSP) of Ascaris suum.";
 RL J. Mol. Biol. 263:284-296(1996).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE=99096891; PubMed=9878374;
 RA Haaf A., Leclaire L. III, Roberts G., Kent H.M., Roberts T.M.,
 RA Stewart M., Neuhaus D.;
 RT "Solution structure of the motile major sperm protein (MSP) of
 RT Ascaris suum - evidence for two manganese binding sites and the
 RT possible role of divalent cations in filament formation.";
 RL J. Mol. Biol. 284:1611-1624(1998).
 CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
 CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
 CC -!- SUBUNIT: FORMS FILAMENTS 10 NM WIDE, WITH A CHARACTERISTIC
 CC SUBSTRUCTURE REPEATING AXIALLY AT 9 NM.
 CC -!- TISSUE SPECIFICITY: SPERM.
 CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15680; AAA29375.1; -.
 DR EMBL; X94249; CAA63933.1; -.
 DR PIR; A45944; A45944.
 DR PDB; 1MSP; 07-DEC-96.
 DR PDB; 3MSP; 20-APR-99.
 DR InterPro; IPR000535; MSP_domain.
 DR Pfam; PF00635; MSP_domain; 1.
 KW Cytoskeleton; Acetylation; Sperm; Multigene family; 3D-structure.
 FT INIT_MET 0 0
 FT MOD_RES 1 1 ACETYLATION.
 FT CONFLICT 113 113 G -> D (IN REF. 1).
 SQ SEQUENCE 126 AA; 14259 MW; 477DCEF6F4CFDD8F CRC64;

 Query Match 18.3%; Score 23; DB 1; Length 126;
 Best Local Similarity 100.0%; Pred. No. 2.1e-17;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 103 QFRREWFQDGMVRRKNLPIEYN 125
 |||||
 DB 103 QFRREWFQDGMVRRKNLPIEYN 125
 |||||

 RESULT 8
 MSP1_ONCVO STANDARD; PRT; 126 AA.
 ID MSP1_ONCVO
 AC P13262;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Major sperm protein 1 (MSP1).
 OS Onchocerca volvulus.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Onchocerca.
 OX NCBI_TaxID=6282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89365002; PubMed=2770787;


```
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 1.
GN MSP-1.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROI;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. Forms an extensive filament system that extends
CC from sperm villipoda, along the leading edge of the pseudopod (by
CC similarity).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24499; AAA29146.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 125 AA; 13858 MW; 0255C31F187549BC CRC64;

Query Match 15.1%; Score 19; DB 1; Length 125;
Best Local Similarity 100.0%; Pred.No. 4.le-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 EWFQGDGMVRRKNLPIEYN 125
Db 106 EWFQGDGMVRRKNLPIEYN 124

RESULT 12
MSP2_GLORO
ID MSP2_GLORO STANDARD; PRT; 125 AA.
AC P53022;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 2.
GN MSP-2.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROI;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. Forms an extensive filament system that extends
CC from sperm villipoda, along the leading edge of the pseudopod (by
CC similarity).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24499; AAA29146.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 125 AA; 13858 MW; 0255C31F187549BC CRC64;

Query Match 15.1%; Score 19; DB 1; Length 125;
Best Local Similarity 100.0%; Pred.No. 4.le-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 EWFQGDGMVRRKNLPIEYN 125
Db 106 EWFQGDGMVRRKNLPIEYN 124

RESULT 13
MSP3_GLORO
ID MSP3_GLORO STANDARD; PRT; 125 AA.
AC P53023;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 3.
GN MSP-3.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROI;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. Forms an extensive filament system that extends
CC from sperm villipoda, along the leading edge of the pseudopod (by
CC similarity).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24500; AAA29147.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 125 AA; 13801 MW; 1B04AD5756511B13 CRC64;
```

```
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24500; AAA29147.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 125 AA; 13801 MW; 1B04AD5756511B13 CRC64;

Query Match 15.1%; Score 19; DB 1; Length 125;
Best Local Similarity 100.0%; Pred.No. 4.le-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 EWFQGDGMVRRKNLPIEYN 125
Db 106 EWFQGDGMVRRKNLPIEYN 124

RESULT 13
MSP3_GLORO
ID MSP3_GLORO STANDARD; PRT; 125 AA.
AC P53023;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 3.
GN MSP-3.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROI;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm crawling. Forms an extensive filament system that extends
CC from sperm villipoda, along the leading edge of the pseudopod (by
CC similarity).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24501; AAA29148.1; -.
CC HSSP; P27439; IMSP.
CC InterPro; IPR000535; MSP_domain.
CC Pfam; PF00635; MSP_domain; 1.
CC Cytoskeleton; Acetylation; Sperm; Multigene family.
CC INIT_MET 0 0 BY SIMILARITY.
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 125 AA; 13844 MW; 0255C5A9C31E49BC CRC64;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2002, 11:01:31 ; Search time 24.47 Seconds
(without alignments)
890.779 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 126
Sequence: 1 AQSVPGGDIQTQPGTKIVEN.....EWFQGDGMVRRKNLPIEYNP 126

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match %	Description
1	126	100.0	Q19832 caenorhabdi
2	101	80.2	Q9tvw5 caenorhabdi
3	101	80.2	Q9n2m1 caenorhabdi
4	100	79.4	Q23519 caenorhabdi
5	90	71.4	Q94053 caenorhabdi
6	86	68.3	Q27280 caenorhabdi
7	71	56.3	Q21244 caenorhabdi
8	59	46.8	Q17856 caenorhabdi
9	53	42.1	Q18461 caenorhabdi
10	52	41.3	Q95PJ7 caenorhabdi
11	52	41.3	Q95XU7 caenorhabdi
12	51	40.5	Q23524 caenorhabdi
13	22	17.5	Q9GNW0 onchocerca
14	22	17.5	Q9GNV9 onchocerca
15	22	17.5	Q9gnv8 onchocerca
16	19	15.1	Q23428 caenorhabdi

17	18	14.3	107	5	Q9GNW2	Q9gnw2 onchocerca
18	17	13.5	107	5	Q9GNI0	Q9gni0 mansonella
19	17	13.5	107	5	Q9GNW1	Q9gnw1 onchocerca
20	17	13.5	107	5	Q9GNX1	Q9gnx1 mansonella
21	17	13.5	107	5	Q9GNX0	Q9gnx0 mansonella
22	17	13.5	107	5	Q9GNW9	Q9gnw9 mansonella
23	17	13.5	107	5	Q9GNW8	Q9gnw8 mansonella
24	17	13.5	107	5	Q9GNW7	Q9gnw7 mansonella
25	17	13.5	141	5	Q26316	Q26316 dictyocaulu
26	16	12.7	107	5	Q9GNW6	Q9gnw6 mansonella
27	15	11.9	107	5	Q9GNW5	Q9gnw5 mansonella
28	15	11.9	107	5	Q9GNW4	Q9gnw4 mansonella
29	12	9.5	442	5	Q9NAP2	Q9nap2 caenorhabdi
30	11	8.7	484	5	Q9NAM2	Q9nam2 caenorhabdi
31	9	7.1	99	5	Q9NAF5	Q9naf5 caenorhabdi
32	8	6.3	83	5	Q26112	Q26112 pratylenchu
33	8	6.3	83	5	Q26096	Q26096 pratylenchu
34	8	6.3	84	5	Q26111	Q26111 pratylenchu
35	8	6.3	85	5	Q26097	Q26097 pratylenchu
36	8	6.3	88	5	Q27405	Q27405 pratylenchu
37	8	6.3	90	5	Q26098	Q26098 pratylenchu
38	8	6.3	108	13	Q9I8H1	Q9i8h1 oncorhynchu
39	8	6.3	351	11	Q91YY8	Q91yy8 mus musculu
40	7	5.6	37	5	Q18145	Q18145 caenorhabdi
41	7	5.6	161	4	Q9NWC0	Q9nwc0 homo sapien
42	7	5.6	167	16	O53195	O53195 mycobacteri
43	7	5.6	175	10	Q93VA3	Q93va3 arabidopsis
44	7	5.6	203	16	Q92FT8	Q92ft8 listeria in
45	7	5.6	231	16	Q9HUP0	Q9hup0 pseudomonas

ALIGNMENTS

RESULT 1

Q19832

ID Q19832 PRELIMINARY; PRT; 127 AA.

AC Q19832;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 14.2 KDA PROTEIN.

GN F26G1.7.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Chissoe S.;

RT "The sequence of C. elegans cosmid F26G1.";

RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RT "Direct Submission.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; U23519; AAK31504.1; -.

DR HSSP; P27439; 3MSP.

DR InterPro; IPR000535; MSP_domain.

DR Pfam; PF00635; MSP_domain; 1.

KW Hypothetical protein.

SQ SEQUENCE 127 AA; 14237 MW; 71671F31BEA5B147 CRC64;

RT "Direct Submission.";
RL EMBL; U29380; AAA68739.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14239 MW; 69F6962DCEACC152 CRC64;

Query Match 79.4%; Score 100; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 7e-100;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGV 61
Db 3 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCGV 62

QY 62 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 63 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 5
Q94053 PRELIMINARY; PRT; 127 AA.
AC Q94053;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T13F2.11 PROTEIN.
GN T13F2.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81122; CAB03362.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 127 AA; 14223 MW; 69F69622069E2147 CRC64;

Query Match 71.4%; Score 90; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 4.4e-89;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61

QY 61 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIT 90
Db 62 VLDPKEAVLLAVSCDAFAFGQEDTNNDRIT 91

RESULT 6
Q27280 PRELIMINARY; PRT; 127 AA.
ID Q27280
AC Q27280;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE C04G2.4 PROTEIN.
GN C04G2.4 OR ZK1251.6 OR K07F5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Hemby C.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA McMurray A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Hemby C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBBJ databases.
DR EMBL; Z70718; CAA94674.1; -.
DR EMBL; Z68222; CAA92502.1; -.
DR EMBL; Z70284; CAA94278.1; -.
DR EMBL; Z70284; CAA94283.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 127 AA; 14236 MW; E5B96631BEBF1419 CRC64;

Query Match 68.3%; Score 86; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 9.1e-85;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 KIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCVLDPKEAVLLAVSCD 75
Db 17 KIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCVLDPKEAVLLAVSCD 76

QY 76 AFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 77 AFAFGQEDTNNDRITVEWTNTPDGAA 102

RESULT 7
Q21244 PRELIMINARY; PRT; 127 AA.
ID Q21244
AC Q21244;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN K05F1.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.

```
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wohldmann P.;
RT "The sequence of C. elegans cosmid K05F1.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U29377; AAA68714.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14228 MW; 7D5D628EF5511AA7 CRC64;

Query Match 56.3%; Score 71; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.4e-68;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 IKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAVLLAVSCDAFAFGQEDTNNDRI 90
Db |||||
32 IKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAVLLAVSCDAFAFGQEDTNNDRI 91

QY 91 VEWNTPTDGAA 101
Db |||||
92 VEWNTPTDGAA 102

RESULT 8
Q17856 PRELIMINARY; PRT; 127 AA.
ID Q17856;
AC Q17856;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C09B9.6 PROTEIN (HYPOTHETICAL PROTEIN R13H9.2).
GN C09B9.6 OR R13H9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bradshaw H.;
```

```
RT "The sequence of C. elegans cosmid C09B9.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA R. Wilson, Bradshaw H.;
RT "The sequence of C. elegans cosmid R13H9.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U50069; AAB37555.1; -.
DR EMBL; AC006681; AAK85492.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 127 AA; 14251 MW; 28E69731A2D5AD2B CRC64;

Query Match 46.8%; Score 59; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.2e-55;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 YGIKTTNMKRLGVDPGVLDPKEAVLLAVSCDAFAFGQEDTNNDRI 101
Db |||||
44 YGIKTTNMKRLGVDPGVLDPKEAVLLAVSCDAFAFGQEDTNNDRI 102

RESULT 9
Q18461 PRELIMINARY; PRT; 127 AA.
ID Q18461
AC Q18461;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN C34F11.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bentley D.;
RT "The sequence of C. elegans cosmid C34F11.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U46753; AAA85759.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14221 MW; 53D359DB8EA5AB74 CRC64;

Query Match 42.1%; Score 53; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 3.7e-49;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 NMKRLGVDPGCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
|
Db 50 NMKRLGVDPGCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 102
|

RESULT 10
Q95PJ7 PRELIMINARY; PRT; 77 AA.
AC Q95PJ7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F09C12.7 PROTEIN.
GN F09C12.7.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

RA [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Anderson K.;

RT "The sequence of C. elegans cosmid F09C12.";

RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.

DR EMBL; U28929; AAA68346.1; -.

SQ SEQUENCE 77 AA; 8697 MW; A3B5FFBDA6FA8F81 CRC64;

Query Match 41.3%; Score 52; DB 5; Length 77;
Best Local Similarity 100.0%; Pred. No. 2.9e-48;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 MKRLGVDPGCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
|
Db 1 MKRLGVDPGCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 52
|

RA Hallsworth K.;
RT "The sequence of C. elegans cosmid ZK546.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29380; AAA68736.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 133 AA; 14899 MW; B283C4CDFDF9B76A CRC64;

Query Match 40.5%; Score 51; DB 5; Length 133;
Best Local Similarity 100.0%; Pred. No. 5.6e-47;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 KRLGVDPPCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 101
Db 58 KRLGVDPPCGVLDPKAEVLLAVSCDAFAFGQEDTNNDRITVEWTNTPDGAA 108

RESULT 13
Q9GNW0 PRELIMINARY; PRT; 107 AA.
AC Q9GNW0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAJOR SPERM PROTEIN (FRAGMENT).
GN MSP.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVNODBRA2C3 ; PubMed=11428336;
RX MEDLINE=21321180; Post R.J.;
RA Morales Hojas R., Post R.J.;
RT "Regional genetic variation in the major sperm protein genes of
RT Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";
RL Int. J. Parasitol. 30:1459-1465(2000).
DR EMBL; AJ404206; CAC20740.1; -.
DR HSSP; P27439; IMSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 12018 MW; 0C5E932130E5DF07 CRC64;

Query Match 17.5%; Score 22; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 9.2e-16;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 IKTNTMKRLGVDPPCGVLDPKE 66
Db 39 IKTNTMKRLGVDPPCGVLDPKE 60

RESULT 14
Q9GNV9 PRELIMINARY; PRT; 107 AA.
AC Q9GNV9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAJOR SPERM PROTEIN (FRAGMENT).

GN MSP.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVNODBRA2C4 ; PubMed=11428336;
RX MEDLINE=21321180; Post R.J.;
RA Morales Hojas R., Post R.J.;
RT "Regional genetic variation in the major sperm protein genes of
RT Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";
RL Int. J. Parasitol. 30:1459-1465(2000).
DR EMBL; AJ404207; CAC20741.1; -.
DR HSSP; P27439; IMSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 12017 MW; 0C5E932C86E5DF07 CRC64;

Query Match 17.5%; Score 22; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 9.2e-16;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 IKTNTMKRLGVDPPCGVLDPKE 66
Db 39 IKTNTMKRLGVDPPCGVLDPKE 60

RESULT 15
Q9GNV8 PRELIMINARY; PRT; 107 AA.
AC Q9GNV8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAJOR SPERM PROTEIN (FRAGMENT).
GN MSP.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVNODBRA2C5 ; PubMed=11428336;
RX MEDLINE=21321180; Post R.J.;
RA Morales Hojas R., Post R.J.;
RT "Regional genetic variation in the major sperm protein genes of
RT Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";
RL Int. J. Parasitol. 30:1459-1465(2000).
DR EMBL; AJ404208; CAC20742.1; -.
DR HSSP; P27439; IMSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 11987 MW; 0C4325F7F6E5DF07 CRC64;

Query Match 17.5%; Score 22; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 9.2e-16;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 IKTNTMKRLGVDPPCGVLDPKE 66
Db 39 IKTNTMKRLGVDPPCGVLDPKE 60

Search completed: September 24, 2002, 11:04:07
Job time: 156 sec

